DOCUMENT RESUME

ED 061 790 FL 002 948

TITLE	Linguistics: A Bibliography of Selected Rand Publications.
INSTITUTION	Rand Corp., Santa Monica, Calif.
REPORT NO	SB-1035
PUB DATE	Jul 71
NOTE	23p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Annotated Bibliographies; *Computational Linguistics; *Computers; Cybernetics; Grammar; Information Processing; Information Theory; Language Development; *Language Research; Linguistic Performance; Linguistics; *Linguistic Theory; Machine Translation; Russian; Semantics; Structural Analysis; Syntax; Word Frequency

ABSTRACT

This annotated bibliography lists approximately 100 books, reports, Rand memoranda, and papers, written between 1961 and 1971, in the fields of linguistic theory, language research, computational linguistics, and formal linguistics. Many documents involve the use of computers in problems of language analysis; many concern the analysis of Russian. Several works on machine translation are also listed. (VM)

ED 061790

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

A BIBLIOGRAPHY OF

Selected Rand Publications

The Rand Corporation
1700 Main Street
Santa Monica
California 90406

July 1971
SB-1035

LINGUISTICS

FL 002 948

BOOKS

Book Exact Methods in Linguistic Research, by O. S. Akhmanova, I. A. Mel'chuk, R. M. Frumkina, E. V. Paducheva. Translated from the Russian by D. G. Hays, D. V. Mohr. 1963.
An attempt to illuminate several scientific results in the application of exact methods to linguistic research. Often termed "mathematical linguistics," the essence of this new approach and its content consist not of creating some special kind of "linguistics," but rather of perfecting, of making accurate, reliable, and modern the methods of conventional linguistic research. This book deals with: (a) questions of general linguistics which relate the following discussions of concrete methods of exact study and descriptions of linguistic phenomena to earlier linguistics; (b) the place and role of machine translation in contemporary linguistics in both a theoretical and a practical sense; (c) possible applications of statistical methods to linguistic research, together with a discussion of basic principles and concepts of statistical analysis; (d) possible applications of information theory to language study. The book does not treat linguistic applications of "non-quantitative mathematics--in particular, mathematical logic. This work was translated in order to show the direction and scope of the best Russian research in modern linguistics. 194 pp. Bibliog. (Published by the University of California Press, 1963, $6.00. Available only from booksellers or the publisher.) (Also published as R-397-PR, September 1963.)

Book Russian Derivational Dictionary. D. S. Worth, A. S. Kozak, D. B. Johnson. September 1970.
A unique tool for Russian language studies: the complete vocabulary of more than 100,000 words from the 1959 Ozegov and Sapiro dictionary, segmented into their constituent elements and grouped into 10,953 families ("derivational nests"). Each root has 1 to 16 variant subfamilies, alphabetically arranged by root, occurrence root, preroot segment, and postroot segment. No such derivational dictionary exists for any other modern language. This is the first for Russian since 1836, despite the central importance of derivational morphology to the Russian linguistic system. Massive theoretical and practical difficulties arise because, for example, Russian phonemes may participate in more than one series of letter alternations, and most compounds are also suffixal derivatives. Each such case was considered

separately for maximum usefulness of grouping. The clerical part of the 7-year effort was done by computer, with manual changes in about 30% of the cases. 771 pp. Ref. (MW) (Published by American Elsevier Publishing Co., $22.95. Available only from booksellers or the publisher.)

REPORT

R-661-PR The Relational Data File and the Decision Problem for Classes of Proper Formulas. R. A. DiPaola. February 1971.
In connection with Rand's Relational Data File, the class of proper formulas has been proposed as comprising those formalizations of questions to be processed by the RDF that are especially suitable for machine processing (see R-511 and RM-5428). The subclasses of proper formulas depend on the identity of the logical primitives employed. Different sets of primitives give rise to different classes of proper formulas. In this report it is shown that the decision problem is solvable for (1) the class of proper prenex formulas (on any set of primitives), (2) the class of proper formulas on negation, disjunction, and existential quantification, (3) the class of proper formulas on negation, implication, and existential quantification, and (4) the class of proper formulas on negation, disjunction, implication, and existential quantification. Thus, for each of these classes there is a mechanical decision procedure that determines if an arbitrary formula (on the relevant set of primitives) is a member of the class. 26 pp. Ref. (Author)

RAND MEMORANDA

RM-2712-AFOSR On the Value of a Dependency Connection. D. G. Hays. 20 January 1961.
The first of a series of publications dealing with transformation and word classification. This Memorandum provides a computational aid for linguistic analysis. Beginning with a definition of value for certain classes of dependency types, the study develops a procedure for assigning estimated values to words in a text. 26 pp. (See also RM-2713, RM-2714.)

RM-2713-AFOSR Procedures for the Determination of Distributional Classes. K. E. Harper. 23 January 1961.
The second in a series of publications dealing with transformation and word classification. This Memorandum describes typical procedures to be used in the automatic analysis of a Russian physics text and discusses some of the problems involved. Data derived from semiautomatic analysis are also presented. 24 pp. (See also RM-2712, RM-2714.)

RM-2714-AFOSR Transformational Criteria for the Classification of Predicative Genitive Constructions in Russian. D. S. Worth. 24 January 1961.
The third in a series of publications dealing with transformation and word classification. This Memorandum analyzes Russian constructions containing genitive substantives that must be modified by another unit, the exact form of which is not relevant. The transformationally determined varieties of such constructions are catalogued, and it is shown that all such constructions contain either an actual or an implied predication. 19 pp. (See also RM-2712, RM-2713.)

RM-2803-AFOSR Six Tasks in Computational Linguistics. K. E. Harper, D. G. Hays, D. S. Worth, T. W. Ziehe. October 1961.
A description of six task areas of Rand work in computational linguistics. These areas include distributional semantics, grammatical transformations, derivational-family classification of Russian words, text processing, glossary editing, and text improvement. 114 pp.

RM-3161-PR Suprasyntactics. D. S. Worth. June 1962.
Part of a continuing study of Russian and English syntactic structure. The Memorandum proposes an explanatory model of the "suprasyntactic" structures of negation, interrogation, and, especially, emphasis. This last category is treated as a formal linguistic operation parallel to negation and interrogation. All three categories are best described as operations ("transformation") performed upon the various nodes and terminal units in a phase-structure (immediate constituent) tree from the entire sentence down to submorphemic sound sequences. Illustrative examples are taken from English, Russian, German, French, and Czech. 21 pp.

RM-3190-PR Two Operators for Determining Agreement (for Automatic Syntactic Analysis) by I. A. Mel'chuk. Translated by D. V. Mohr. June 1962.
A translation of one of the more interesting recent (1961) Soviet publications dealing with structural and applied linguistics. This is one of several areas of language research relating to Rand's study of automatic language-data processing. Two interrelated programs are presented for the automatic analysis of Russian sentences, or of sentences in any language characterized by grammatic agreement among certain word classes that must be tested to derive the proper syntactic structure for each sentence being processed. 56 pp.

RM-3235-PR Studies in Russian Morphology--I: the Suffix -aga. D. S. Worth. August 1962.
An analysis of the formal and semantic structure of Russian nouns derived from nouns, adjectives, and verbs by means of the suffix -aga. The formal devices of derivation include not only addition of the suffix itself, but also specific procedures for selecting a derivational stem and specific consonantal and vocalic alternations occasioned by the suffix. The semantic differential between base form and derivative in -aga is described in terms of differences of referential meaning and of evaluative and distributional markings. The parallelism of formal and nonformal procedures is indicated, as well as the interaction of derived and nonderived -aga forms. 45 pp.

RM-3339-PR Preliminary Codes and Rules for the Automatic Parsing of English. J. J. Robinson. December 1962.
An interim report on automatic computer analysis (parsing) of English intended to lead to techniques that will enable the electronic computer to receive natural English sentences and convert them internally to formal structures suitable for further machine manipulation. The Memorandum describes what is essentially a two-part machine grammer. The first part develops a system of codes for indicating the syntactic role or roles a particular word plays. This is the parts-of-speech category of the grammar. The second part consists of a set of parsing rules for listing all the grammatically permissible combinations of adjacent codes and supplying a new code for each combination. This permits the machine to determine how words have been put together in substructures (subject, verb, object, etc.) in actual sentences. 149 pp.

RM-3383-PR High-Frequency Words and Occurrence Forms in Russian Physics. A. S. Kozak. October 1962.
One of a series of studies on the linguistic analysis of scientific Russian text. Rand Memorandum RM-2655, A Glossary of Russian Physics, tabulates the separate word forms appearing in 265,417 running words of physics text. This Memorandum lists the 1000 highest-frequency words and the 1000 highest-frequency occurrence words of that glossary. Each list is arranged in descending order of frequency of occurrence and in alphabetical order. Cumulative frequencies and percentage of text covered are shown with each entry. 102 pp. (See also RM-2916-PR.)

RM-3610-1-PR Studies in Machine Translation: Bibliography of Russian Scientific Articles. K. E. Harper, D. G. Hays, A. S. Kozak, B. J. Scott. November 1966.
A catalog of 702 Russian-language articles currently available on magnetic tape at Rand. The articles were published during 1953-59 in leading Soviet scientific journals and books in the fields of physics, mathematics, astrobotany, cybernetics, and physiology. The tape was prepared according to the codes and formats described in RM-2065-1, RM-4390-PR, and RM-4540-PR and is available with or without syntactic annotations. Each entry contains the transliterated title and its English translation; author's name transliterated by Rand rules; the usual bibliographic identification; and the number of running words. When possible, English titles were taken from the Doklady Akademia Nauk SSR and Soviet Physics. This Memorandum supersedes RM-2069 and RM-3610-PR. 125 pp.

RM-3625-PR The Position of Prepositional Phrases in Russian. K. E. Harper. May 1963.
A study of the correct structuring of prepositional phrases in sentences, a problem frequently encountered in the automatic parsing of Russian texts. Studies of text samples indicate that, when other criteria are absent, the syntactic governors of prepositions can be determined with a high degree of accuracy by reference to the relative position and part of speech of elements in the clause. 24 pp.

RM-3774-PR Studies of Predication in Russian--I. Predicative Case, Short Form Adjectives and Predicatives. H. Birnbaum. January 1964.

The first, theoretical part of the series of studies on certain aspects of predication (predicative function) and its formal expression in Contemporary Standard Russian (CSR). This series is primarily concerned with such instances where, in Russian, the predicate is denoted by means other than a finite verb, i.e., by infinite verbal forms (infinitives and gerunds) as well as nonverbal forms, including the zero-morpheme of the copula verb. In its sequels the series makes use of automatic and semi-automatic procedures applied on corpora of analyzed Russian physics texts available at the Rand Linguistic Research Project, in addition to evidence from other Russian linguistic materials. 86 pp.

RM-3889-NSF Report of a Summer Seminar on Computational Linguistics. D. G. Hays. February 1964.
A report on the Seminar on Computational Linguistics held at RAND 8 July- 30 August 1963. The staff taught elementary programming, linguistic theory, mathematical linguistics, and existing programs for linguistic operations to twenty professors of linguistics. The Memorandum includes a detailed schedule, an analysis and evaluation of the seminar by the participants, and a list of all persons involved. The evaluation indicates that most participants, while attempting to obtain further education in this area, will meantime begin using computers in their own research, and that many of them will also teach this subject--some in special courses, some in general linguistics courses, and others only to research students. 64 pp.

RM-3892-PR Linguistic and Non-Linguistic "Understanding" of Linguistic Tokens. D. M. MacKay. March 1964.
An examination of the difference between recognizing an utterance as a "symptom" of the state of affairs in or confronting its originator and recognizing it as a linguistic tool. It is suggested that, although only the first kind of understanding is required to enable a computer to accept data and answer questions in verbal form, such ability does not guarantee its comprehension of all aspects of language. To reach full linguistic compenhension, the program must also embody at least a skeleton representation of the linguistic context of an utterance. By use of such a model, a computer should be able to meet more sensitive tests of linguistic comprehension. 19 pp.

RM-3894-3 Annotated Bibliography of RAND Publications in Computational Linguistics. D. G. Hays, B. Henisz-Dostert, M. L. Rapp. June 1967.
A listing, with summaries, of 143 publications by RAND authors on linguistic theory (especially dependency theory), research methods, application of computers, the English and (especially) Russian languages, information retrieval (especially the Relational Data File), automatic content analysis, psycholinguistics, and character readers (pattern recognition). Only a few of the RAND studies of artificial intelligence (neural nets, automata) are included, and information theory is excluded, in its usual sense of probabilistic analysis of communication channels. There are 29 new entries and one revision. Citations are arranged by subject and indexed by author and RAND publication number. 45 pp.

RM-3979-PR Studies in Information Processing Theory: Similarity and Familiarity in Verbal Learning. H. A. Simon, E. A. Feigenbaum. February 1964.
A presentation of results obtained by simulating various verbal learning experiments with the Elementary Perceiving and Memorizing Program (EPAM), an information processing theory of verbal learning. Predictions were generated for experiments manipulating intra-list similarity (Underwood), inter-list similarity (Bruce), and familiarity and meaningfulness. The stimulus materials were nonsense syllables, learned in paired-associate fashion. The predictions made by the model are generally in good agreement with the experimental data. The fit of the EPAM predictions to the Chenzoff data is particularly significant since it lends support to the hypothesis that the mechanism by which a high degree of meaningfulness of items facilitates learning is the high familiarity of these items. 38 pp.

RM-4005-PR Automatic Parsing and Fact Retrieval: A Comment on Grammar, Paraphrase, and Meaning. J. J. Robinson. February 1964.
A consideration of the development of an automatic parsing grammar as part of a fact retrieval system. The aim of automated parsing (machine analysis) of sentences is to enable an electronic computer to manipulate and process information received in the form of natural English sentences. This Memorandum contends that, through the refinement of parsing grammers, insights can be gained into the nature of meaning, for example, in the creating of a paraphrase. These insights can then be

used in the development of fact retrieval systems receiving requests and giving responses in natural English. 60 pp. Bibliog. (See also RM-3339-PR.)

RM-4065-PR Computers and comprehension. M. Kochen, D. M. MacKay, M. E. Maron, M. Scriven, L. Uhr. April 1964.
An examination of the problem of understanding language by an antifact, from a four-week symposium held at Rand in the summer of 1963, (one of a series of symposia on mathematical biology). Understanding has both linguistic and nonlinguistic aspects. An organism (natural or artificial) may be said to understand adequately some feature of its environment if its internal organizing system can take adequate account of it (without necessarily being able to express the understanding linguistically). In the linguistic context, a new set of capacities is needed: (1) the development and use of internal models as such; (2) the distinction between linguistic and direct experiential input; and (3) the capacity to learn. For understanding, linguistic or not, the information system must have something equivalent to an internal model of what is understood. To understand the difference between a description and a mere symptom of a state of affairs, the information system must have something equivalent to an internal model of the communication process. Four levels of demands for language-handling ability are listed and explained. 71 pp. Bibliog.

RM-4087-PR Dependency Theory: A Formalism and Some Observations. D. G. Hays. July 1964.
A general survey of dependency theory from the viewpoint of the professional linguist, examining in detail the motivation as well as the structure of the theory. Dependency grammars characterize the class of context-free languages, assigning to each sentence of a characterized language a tree structure with minimal syntactic units at the nodes. Transformational or stratified linguistic systems can be constructed on dependency theory. Semantic and psychological considerations motivate specific features of the theory, but are unnecessary as justifications. 46 pp.

RM-4136-PR Language Identification in the Limit. E. M. Gold. July 1964.
A consideration of the problem of identifying a language of some specified class, a language being defined as a subset of the strings on some finite alphabet. Information about the language may be given by a text, an unending succession of examples of strings in the language, or by informant, a source that tells whether successive strings are in the language. Decision procedures are considered which generate successive guesses as to the identity of the language. The language is said to be identified in the limit if these guesses are the same and correct after some finite time. Six variations each of the two basic identification situations are considered. The purpose of the Memorandum is to apply the methods and results of limiting recursion to the problem of identifying a language on the basis of an endless flow of information, presented in a manner intended to model the circumstances under which a human learns a language. 40 pp.

RM-4224-PR The Logic of Cognate Recognition in Historical Linguistics. M. Kay. September 1964.
A formalization of one of the steps in the comparative method in historical linguistics; i.e., recognition of modern derivatives of prehistoric phonemes. The theory assumes that words of a hypothetical prehistoric language should be constructed so as to minimize the total number of phonemes in the language and the total number of statements needed to account for the forms of modern words. It provides an algorithm for a computer program, but even small increases in the amount of data to be considered exhaust present computer techniques. The theory nevertheless provides a basis for more efficient heuristic procedures. 26 pp.

RM-4283-PR A Parsing Program for Categorial Grammars. M. Kay. September 1964.
A description of a parsing procedure, written for the IBM 7090 computer, which finds all the structures of a sentence allowed by a categorical grammar. It achieves high speed and economy of storage by first putting grammatical tags into parenthesis-free form and then compiling them into lists of computer instructions. 32 pp.

RM-4383-PR Prepositional Phrases and Automatic Parsing. K. E. Harper. February 1965.
Presentation of a technique for making determinable the structure of a Russian prepositional phrase preceded by a noun (P-1=N). Along with the author's earlier work on structuring prepositional phrases

in Russian sentences, the Memorandum presents the basic information required to write a computer program to handle the problem of structuring prepositional phrases. 34 pp. Bibliog. (See also RM-2538, RM-3625.)

RM-4390-PR Natural Language in Computer Form. M. Kay, T. W. Ziehe. February 1965.
Description of a scheme for recording text in computer-useable form so that all meaningful typographical distinctions are represented in a standard format. Provision is made for texts in different languages and different alphabets, and for subsidiary material such as parallel translations and comments of interest to users and librarians. The basic set of encoding conventions is indefinitely extensible to accommodate new kinds of material. In particular, the Memorandum: (1) provides special facilities for very large bodies of data (by embedding the text encoding scheme in a general file maintenance system), and a means of correcting and revising filed material; (2) describes computer programs which simplify conversion of text from various sources into the standard format: (3) discusses the problem of and describes a flexible program for printing text which has been recorded in standard format. 90 pp.

RM-4477-PR Studies on Predication in Russian--II. On the Predicative Use of the Russian Infinitive. H. Birnbaum. December 1965.
The second part of a study on the complex function of predication and its formal expression in Contemporary Standard Russian (CSR). This memorandum discusses the predicative use of the Russian infinitive against the background of its entire range of functions (the "semantic spectrum"). The author surveys the predicative and semi-predicative use of the Russian infinitive; studies its peculiarities by applying certain transformational procedures; and discusses how it expresses predicative categories of time, modality, person, aspect, mode of action, etc. Some suggestions for automatically translating Russian predicative infinitives are included. 258 pp. Bibliog. (See also Part I, RM-3774-PR.)

RM-4479-PR An Annotated Bibliography of Publications on Dependency Theory. D. G. Hays. March 1965.
An annotated bibliography of publications on the development of a formal dependency theory in linguistics, the application of the theory to the description of natural languages, and to the design of computer systems for machine translation, information retrieval, and other purposes. 22 pp.

RM-4532-PR Measurement of Similarity Between Nouns. K. E. Harper. May 1965.
A study of distributional semantics, or the classification of words on a behavioral basis, in automatic language data processing. The Memorandum presents some of the relevant behavioral conditions, applies these conditions to specified words in a restricted text, and measures the similarity of behavior of words that are presumed to be similar in meaning. The experiment is intended as a test of the validity of the distributional approach to semantics. 26 pp.

RM-4540-PR The Catalog Input/Output System. M. Kay, F. Valadez, T. W. Ziehe. March 1966.
A complete definition of the format used for catalogs on magnetic tape and a description of the routines of the Catalog Input/Output System. Catalog maps, catalog data, and tape labels are written as logical records in a specially designed blocking format. Beginnings and ends of blocks, physical tapes, and catalogs are explicitly marked. The Catalog Input/Output System offers a variety of input/output unit-control operations in addition to the commands for reading and writing data in catalog format. The reading and writing commands are implemented on three levels: Level III handles individual data; Level II reads and writes logical records; and Level I processes blocks of information. 71 pp. (See also RM-4645-PR.)

RM-4582-PR Complementation in Russian: Theory and Application. A. S. Kozak. September 1965.
A presentation of the concept of complementation in Russian as seen from two points of view. An expansion and modification of the Russian view becomes an hypothesis, given in the form of complementation models. The models serve as a basis for the functional annotation of complements in Russian scientific text within the framework of dependency theory. Structured text in computer form is expected to provide the means to check the

hypothesis and to establish refined syntactic categories. 51 pp. Bibliog.

RM-4645-PR The Catalog: A Flexible Data Structure for Magnetic Tape. M. Kay, T. W. Ziehe. October 1965.
An outline of a generalized storage scheme for large files of highly structured data (or catalogs) and a description of their realization on magnetic tape. Each datum, large or small, is assigned to one of a number of data classes of which a user may define any number. The over-all organization of a catalog is given by a map that imposes a tree structure on the set of data classes. Catalogs may participate as individual data in other catalogs so that the structure of a file may be recursive. A flexible addressing scheme not only facilitates the retrieval of data and sets of data in response to easily formulated requests but also provides a basis for updating catalogs. General procedures called transformations can be used to derive catalogs with new structures from existing catalogs. 32 pp. (See also RM-4390-PR.)

RM-4654-PR (Part I) PARSE: A System for Automatic Syntactic Analysis of English Text--Part I. J. Robinson, S. Marks. September 1965.
A description of PARSE, a system for the automatic syntactic analysis of English sentences. It is capable of analyzing and labeling the structure of a variety of sentences. The system employs a parsing logic which both applies the grammar exhaustively and disengages it from the routines that apply it to text, so that changes in grammar will not affect the program. The Memorandum describes the linguistic features and details the operation of the program. 207 pp. (See also RM-3339-PR, RM-4654-PR, Pt. II.)

RM-4654-PR (Part II) PARSE: A System for Automatic Syntactic Analysis of English Text--Part II. J. Robinson, S. Marks. September 1965.
A listing of the three major components of PARSE, a system for the automatic syntactic analysis of English sentences. It contains: (1) a glossary of the words used, grouped alphabetically by families; (2) a presentation of the words organized by the grammar codes devised for them; and (3) a listing of 26 words which, because of their excessive length, would not fit the format and had to be removed from the main body of the glossary. A final section lists errata. 270 pp. (See also RM-3339-PR, RM-4654-PR, Pt. I.)

RM-4828-PR Studies in Inter-sentence Connection. K. E. Harper. December 1965.
A systematic analysis of portions of the Russian physics text processed at RAND, to determine the extent of repetition in adjacent sentences. Recurrence of words in all pairs of contiguous sentences in the text (2467 pairs) was recorded in a machine printout; sentence pairs for which word recurrence was not automatically recorded were visually inspected for other types of recurrence (through lexical stems, pronouns, synonyms, and paraphrases). The extent of the different types of recurrence is reported, and features of the recurring items are discussed. Sentence pairs characterized by nonrecurrence (12 percent of the total) are examined, and the relevance of inter-sentence recurrence and nonrecurrence to automatic syntactic analysis and abstracting is suggested. 21 pp. (See also RM-3610-PR.)

RM-4920-PR Computer Routines to Read Natural Text in Complex Formats. P. A. Graves, D. G. Hays, M. Kay, T. W. Ziehe. August 1966.
A description of a system of IBM 7040/44 subroutines that will accept natural-language input with complex formats--e.g., from books, journals, questionnaires, clippings, library catalog cards, etc.--prepared by any typesetting device or other machine (typewriter, keypunch, etc.). Inputs are transcribed by the computer into a standard code for machine processing and can be rearranged into any desired format for storage or output. Different kinds of information are recognized by explicit markers, position on the line or page, or syntactic clues given by other items. The subroutines can be used singly or together; they may be called from either FORTRAN or MAP programs. A detailed programmers' guide is included. 143 pp.

RM-4933-PR The Tabular Parser: A Parsing Program for Phrase Structure and Dependency. M. Kay. July 1966.
A comparison of phrase-structure and dependency grammars and description of a computer program for parsing sentences concurrently by both systems with cross references. A simple, concise notation system, rule tables, and directions for punched-card input are given for use with a FORTRAN IV program with MAP subroutines.

The program itself is the same without regard to language, rules, or directory entries input, and is based on a morsel strategy, described in ALGOL logic. 56 pp.

RM-4986-PR Computational Linguistics: Bibliography, 1965. D. G. Hays, B. Henisz-Dostert, M. L. Rapp. April 1966.
A bibliography citing 700 U.S. and foreign articles, reports, and books relevant to the fields of computational linguistics and documentation. A selective coverage of classification theory, computation and programming, computers and hardware, nonnumerical applications of computers, and psycholinguistics is included. In the area of linguistics, the editors have taken a broad view of structural theory and semantics. 79 pp.

RM-5077-PR Some Combinatorial Properties of Russian Nouns. K. E. Harper. September 1966.
An analysis of nouns in approximately 180,000 words of parsed Russian text, in terms of "valence;" that is, the tendency of nouns to occur as the syntactic governor of a noun in the genitive case. The preliminary results of the study suggest that (1) for some nouns, valence may be attributed to certain inherent semantic properties, and (2) for other nouns, valence may be highly correlated with other "contextual" features, both syntactic and lexical. 54 pp. Refs.

RM-5085-PR A Computer System for Inference Execution and Data Retrieval. R. E. Levien M. E. Maron. September 1966.
Describes the Relational Data File, a computer-based data retrieval system now partly operational, capable of storing millions of facts and of retrieving data both directly and through logical inference. A large collection of information about cybernetics research was used as the data base. Facts are stored in the form of sentences in an artificial information language, each representing a binary relationship between two entities. Data are entered by means of special input forms; a translation program in a special programming language, FOREMAN, derives the relational sentences that are then internally coded and stored in the data file. An intensional file that stores facts about relationships permits many data sentences to be derived when needed, rather than stored explicitly. Counting space for all auxiliaries, one IBM 1301 disk holds 232,000 sentences in the information language. The system operates on an IBM 7044 computer. 35 pp. Refs. (See also RM-4793-PR.)

RM-5108-1-PR Russian-English Dictionary of Cybernetics and Computer Technology, Second Edition, Revised and Enlarged. W. B. Holland. February 1969.
The second edition of the Rand Russian-English Dictionary of Cybernetics and Computer Technology contains over 300 new major entries, making a total of 2050 major entries and about 3300 subentries; each subentry defines a key term as it appears in frequently encountered phrases and special constructions. Many corrections and improvements to the previous definitions have been made. As before, entries are made for terms encountered in the Soviet cybernetic literature, without any attempt to define the field or to assemble a complete glossary. The dictionary is entirely machine-processed from punchcards and magnetic tape, and is photoreproduced directly from computer printout. 261 pp. (MW)

RM-5143-PR A Deep Index of Derivational Morphology. D. S. Worth. September 1966.
A discussion of a projected "deep index" to the extant literature on Russian derivational morphology. The Memorandum describes the type of literature being abstracted, the classification of information into discrete categories for retrieval, the steps taken to prepare data for conversion to machine format, and the transliteration and abbreviation conventions used in the index. An example of Russian scholarly text is appended, together with the format in which this text appears after processing. 37 pp.

RM-5150-PR A Simple Scheme for Formalizing Data Retrieval Requests. F. M. Tonge. May 1967.
A description of an experimental scheme for translating data retrieval requests, expressed in reasonable "natural" language, into a formalized format. The simple translation procedure, which immediately feeds its proposed translation back to the analyst, uses rules depending on the position of words in the request and simple semantic information about the class of words. It is programmed in a list processing language (IPL-V) and is oriented toward a relational file organ-

ization. The present scheme is designed to process a single request rather than a dialogue of requests and retrievals. Also, the translation procedure is not yet coupled to a retrieval process. 38 pp. Refs.

RM-5168-PR The Quantification of Functional Load: A Linguistic Problem. C. F. Hockett. October 1966.
An approach to the quantification of the functional load carried by the sounds of a language in distinguishing one utterance from another. By applying algebraic grammar and matroid lattice theory, criteria are obtained which are satisfied by three measures of entropy used by Shannon in information theory (measured in binits, in shannons, and as relative entropy). The method can be used to measure the work of phonemes, allophones, or components, and should therefore be useful regardless of the outcome of present disputes in the field of phonology. Subjective impressions suggest that there is a balance point in speech between the force of least effort, making for lowest redundancy, and the practical need to be understood, pressing for lower entropy. Research is suggested to test this hypothesis and discover whether it applies equally to all speech communities. 33 pp.

RM-5194-RADC A File of Russian Text with Syntactic Annotations. D. G. Hays, D. Gottshall, A. S. Kozak. December 1966.
A description of the syntactical annotations to the Russian language text file catalogued in RM-3610-1-PR. Over a million running words of text include 702 journal articles on physics, mathematics, astrobotany, and physiology, supplemented by syntactic examples from a Russian grammar. Dependency theory notation has been used. The complete text file, with every word annotated, occupies seven reels of magnetic tape, coded for computer use according to the RAND text and catalog system. 28 pp. Refs. (See also RM-4087-PR, RM-4390-PR, RM-4540-PR, RM-4582-PR, RM-5243-RADC, RM-5252-RADC.)

RM-5209-PR The Queens Grammar. S. L. Lieman. January 1967.
The famous eight-queens chess problem extended into the area of linguistics. The problem is presented and then generalized into a language over strings of integers. Various formalisms are shown to be able to describe this language, including arithmetical relations defined on each pair of integers in the string, programs for actually generating and parsing strings, and a relational grammar. It is shown that context-sensitive grammar and transformational grammar are likely to be capable of defining the language. The simplicity and elegance of the arithmetical formulation and the computer programs that are easily derived from it are contrasted to the counterintuitive properties of relational grammars, context-sensitive grammars, and transformational grammars that define the same language. A natural language counterpart to the Queens Language is given, and new measures of linguistic adequacy are defined. The implications of these new measures of adequacy to efforts for finding appropriate linguistic formalisms is discussed. 47 pp. Refs.

RM-5210-PR Linguistic Relativity and the Language Learning Process. R. M. Schwarcz. December 1966.
A five-stage analysis of the language-learning process, and an investigation of whether this analysis supports or contradicts the Whorfian hypothesis of linguistic relativity. The syntactic constructions of a language influence the types of conceptual relationship that the child perceives. Words and phrases grouped together by the conventions of the language have a determining effect on conceptual representations associated with them. This effect is somewhat mitigated by the transformational phase of language learning, which opens up the full range of stylistic devices available in a language, thus extending the number of percepts and concepts that may gain expression in that language. However, earlier perceptual and conceptual habits will probably persist throughout a person's adult life unless he is forced to change by some new experience, such as the mastery of a different language. 13 pp. Refs.

RM-5214-PR Steps Toward a Model of Linguistic Performance: A Preliminary Sketch. R. M. Schwarcz. January 1967.
A model of human linguistic performance for computer implementation, limited to the input-output interactions and internal processes available on today's digital computers. The model is to understand factual information input to it, answer questions, and seek information by asking questions. It need only learn the morphological and syntactic rules, rules of conceptual inference, and semantic correspondences for whatever natural language subset and data base the user trains it to handle. Its "education" is by input of relational

graph-structure representations of conceptual data structures. Basic to the model is the view of language as a continuous learning process, with semantic and syntactic information treated similarly as different phases of processing, in contrast to traditional views. The assumed stages are those outlined in RM-5210-PR. 43 pp. Refs.

RM-5223-PR Studies in Russian Morphology: II. Vowel-Zero Alternations in Derivation. D. S. Worth. May 1967.
A preliminary report analyzing the vowel-zero alternations of contemporary standard Russian, with emphasis on those alternations specific to the derivational (as opposed to the flexional) system. It is shown that there are two types of vowel-zero alternation in the Russian derivational system. In the first type, a flexional-level alternating vowel-zero morphophoneme is stabilized as either a full vowel or as zero in the course of derivation. In the second type, a full vowel in the derivational base alternates with zero in the derived word, or vice versa. The existence of an alternating vowel-zero morphophoneme on the derivational level is suggested as an explanation for certain of these alternations. 38 pp.

RM-5224-PR Some Aspects of the Thematic Organization of the English Clause. M. A. K. Halliday. January 1967.
An analysis of the English clause from the point of view of its structure as a message: its thematic organization. The study is concerned with spoken English only. Recorded conversations were analyzed by rhythm, stress, and tonality, as well as by word order, to show the distinct patterns underlying a number of apparently unrelated contrasts in English grammar. The three principal systems of structuring communication are information, thematization, and identification. Each assigns a structure to the clause in terms of "given--new" (information), "theme--non-theme" (thematization), and "known--unknown" (identification). Information clauses distinguish between given and new phonologically, by the tone pattern. Thematic clauses distinguish between theme and rheme (non-theme) by the sequencing of elements. Identification clauses relate a known to an unknown by stress pattern. Examination of the thematic patterns of the clause not only reveals the resources available to the speaker for the structuring of information, but can also provide further insights into the better-known aspects of English syntax. 28 pp. Refs.

RM-5243-RADC COLLECT: A Program for the Retrieval of Grammatical Information from Annotated Text. M. Kay, T. D. Taft. January 1967.
A description of a system of computer programs to be used for syntactic analyses of the million-word File of Russian Text with Syntactic Annotations described in RM-5194-RADC. The File consists of sentences, each of which is provided with a dependency structure. To use the COLLECT system, the linguist writes a description of the class of sentences he wishes to examine, and COLLECT edits his request, searches the File, and prints a report. The system is specific for the File of Russian Text but is flexible enough to permit modifications of File information, request terminolgy, and report format. It is representative of a class of retrieval programs needed for many file-search tasks. 34 pp. Refs. (See also RM-4390-PR, RM-4540-PR, RM-4807-PR.)

RM-5252-RADC Toward Exploitation of a File of Russian Text with Syntactic Annotations. D. G. Hays, D. S. Worth. January 1967.
The final report of RAND's work in linguistics for Rome Air Development Center, 1965-66. The work included the compilation of a million-word File of Russian Text with Syntactic Annotations, stored on magnetic tape, which can be duplicated for qualified requesters; and the design of a computer program called COLLECT for retrieving data from the File. The annotations, which are based on dependency theory, include not only systematic connections between dependent and governing words, but grammatical functions, indications of negation, pointers to antecedents of pronouns, and special features. Methods of automatic (statistical) classification usable in reducing File data are discussed, and steps toward ambiguity reduction and automatic parsing are described. Tentative research designs for investigating modal constructions and sequential apposition in Russian with computer assistance are outlined. 45 pp. (See also RM-3610-1-PR, RM-5194-RADC, RM-5243-RADC.)

RM-5279-PR A Formal System for the Logical Analysis of Temporal Relationships Between Intervals of Time. R. Mattison. April 1967.

A formal deduction system for the logical analysis of chronological data (such as before, after, and during relationships). The development of the system was motivated by the requirements of current data-retrieval projects, one of which is RAND's Cybernetics Data-Research Project. A mathematical model of intervals of time is constructed. Twenty basic configurations are used to express all possible time relationships, in which the smallest unit is one calendar day. Axioms are given for making inferences using the Tarski first-order predicate calculus with identity (=), one binary predicate constant ( ), meaning wholly before, and two unary operations: min, meaning the beginning of, and max, the end of. Every sentence in this formal language is such that either it or its negation is deducible from the given axioms. Each proper expression is uniquely readable. Semantic and syntactic completeness are demonstrated, the latter by mathematical logic using Kochen's theorems on ultra-products of models. An appendix gives more general axioms that have both the finite intervals of rationals and the finite intervals of reals as models. 40 pp. Refs.

RM-5345-PR Computational Linguistics: Bibliography, 1966. D. G. Hays, B. Henisz-Dostert, M. L. Rapp. April 1967.
A bibliography of 914 numbered items from 21 countries: books, articles, reports, research memoranda, and symposium contributions relevant to the fields of computational linguistics and documentation that became available in 1966. Particular attention is given to classification theory and practice and to computer hardware and programming for nonnumerical applications (language processing). Entries under Documentation cover abstracting, bibliographic analysis, dissemination, indexing, organization, personnel, user relations, and publications systems. Psycholinguistics receives extensive coverage. There are also items on research methods, semantics, structural theory, style, and on the characteristics of 17 languages and language groups. An author index and an organization index are included, the latter divided by country. This is the third annual compilation of the monthly lists prepared by the RAND computational linguistics project for The Finite String (Association for Machine Translation and Computational Linguistics). 89 pp. (See also RM-4523-PR, RM-4986-PR.)

RM-5452-PR Experiments with a Powerful Parser. M. Kay. October 1967.
A description of a sophisticated computer program for the syntactic analysis of natural languages. The study discusses the notation used to write rules and the extent to which these rules can be made to state the same linguistic facts as a transformational grammar. Whereas most existing programs apply context-free phrase-structure grammars, this new program can analyze sentences with context-sensitive grammars and with grammars of a class very similar to transformational grammars. The program, which is written for the IBM 7040/44 computer, is nondeterministic: The various interpretations of an ambiguous sentence are all worked on simultaneously; at no stage does the program develop one interpretation rather than another. If two interpretations differ only in some small part of a partial syntactic structure, then only one complete structure is stored with two versions of the ambiguous part. The unambiguous portion is worked on only once for both interpretations. Although the current version of the program is written in ALGOL, with very little regard for efficiency, the basic algorithm is inherently much more efficient than any of its competitors. 33 pp.

RM-5580-1-PR An Introduction to the Model Theory of First-Order Predicate Logic and a Related Temporal Logic. R. Mattison. June 1969.
The revised analysis of a formal language having greater expressive capabilities than a first-order predicate language. This new, first-order temporal language can express propositions involving tenses, dates, or quantification over temporal entities. The discussion includes description of the first-order predicate language and the notion of bound and free variables. The notions of a deduction, a theorem of logic, and a consistent set of formulas are introduced, and the deduction theorem and a form of the compactness theorem are stated and proved. Detailed comparison shows that the new, first-order temporal language differs from the traditional one in allowing quantification over intervals and the admission of interval constants. 74 pp. Refs. (KB)

RM-5611-1-ARPA APAREL--A Parse-Request Language. R. M. Balzer, D. J. Farber. September 1969.
A description, without listing, of APAREL, a flexible high-level parsing-language extension to PL/I that has been used in a translator for Dataless Programming (RM-

5290), a powerful syntax and function macro system, and an on-line command parser. Parse requests are stated in a BNF-like format with special sequencing rules. On successful completion, an associated piece of PL/I code is executed, using as data the success or failure of the alternative parsings, and as strings the various elements of the parse at all levels. APAREL includes parse-time semantic checks and facilities for multiple input-output streams. The initial implementation on the IBM 360 comprises (1) a preprocessor that converts APAREL programs into legal PL/I programs, and (2) the run-time parser, which is written in assembly language. APAREL provides the pattern-matching capability normally found only in special-purpose languages such as SNOBOL4 and TMG. 38 pp. (MW)

RM-5733-PR Computational Linguistics: Bibliography, 1967. D. G. Hays, B. Henisz-Dostert, J. I. Houston. July 1968.
A selective, unannotated listing of 664 U.S. and foreign publications that became available in 1967 relevant to the fields of computational linguistics and documentation. The listing includes books, articles, reports, and conference papers. There is selective coverage of classification techniques, computation and programming, with some attention to automata theory, computers and hardware, and social science text-processing by computer. The Documentation heading covers bibliographies, library catalogs, information systems, and publication systems. Computational Linguistics is divided into decoding and encoding; glossaries; machine translation; parsing; synthesis; and text files. Linguistics is subdivided into specific languages, especially English and Russian; algebraic syntax; research methods; and semantics. Arrangement is by subject, with indexes by author and by authors' organizational affiliation (if given). The index by countries includes 109 U.S. organizations, including 14 Federal Government agencies, and 87 organizations in other countries. Previous bibliographies were published as RM-4523-PR, RM-4986-PR, RM-5345-PR. 68 pp.

RM-6053-PR A Directed Random Paragraph Generator. K. E. Harper, S. Y. W. Su. July 1969.
A description of a computer program for generating strings of sentences that have the properties of "real" discourse. The program implements certain linguistic assumptions about "development" and "cohesion" among sentences. The generator consists of two major parts: a data conversion program for establishing compact data tables with associative links and a generation program for synthesizing sentences. The generator, currently operational, is written in PL/I and is run on the IBM 360/65. The objective is to gain a better theoretical understanding of the interrelationship of sentences in a connected discourse. 44 pp. Bibliog. (MJP)

RM-6223-PR Computational Linguistics: Bibliography, 1968. B. Dostert, J. I. Houston, D. V. Lofgren. January 1970.
An unannotated, international bibliography of 573 books, articles, and reports relevant to computational linguistics, with selective coverage in the fields of computation and programming, and social science uses of computers as language processors. In the area of linguistics, structural theory and semantics are viewed rather broadly without being exhaustive. Psycholinguistics and sociolinguistics are selectively covered, as are the structural properties of some languages, especially English and Russian. This bibliography was compiled by scanning 31 journals and the flow of semi-published reports available to the Rand linguistics group. A substantial part of this bibliography has been included in The Finite String, the monthly newsletter of the Association for Computational Linguistics. This Memorandum is the last of the series that includes RM-4523, RM-4986, RM-5345, and RM-5733. The listing is by subject, with an author index. 65 pp. (MW)

RM-6265/1-PR The MIND System: A Grammar-Rule Language. R. M. Kaplan. April 1970.
Description of the MIND Grammar-Rule Language (MGRL), including parsing procedure and grammar rules. MGRL is the formalism for the grammar input to the MIND syntactic analysis component, currently being programmed for the Rand IBM 360/65. MIND (Management of Information through Natural Discourse) is a system consisting of nested and chained modules of high-level programming language statements designed to meet the need for fast and accurate information transactions between senior command, control, and policymaking personnel of the Air Force and other services, and large, heterogeneous, loosely formatted information banks. 80 pp. Bibliog. (TC)

RM-6265/2-PR The MIND System: The Morphological-Analysis Program. M.

Kay, G. R. Martins. April 1970.
A detailed description of the ANALYZE module of the MIND information system. This part of the system receives English sentences as input, segments them into words, and reduces the words to their ultimate lexical components through morphological analysis. Dictionary entries providing morphological, syntactic, and semantic information about the components are retrieved and amalgamated into word-level grammatical representations, and are then concatenated to form a bottom-level grammatical representation of the input sentence. This representation is passed to the PARSE module for further processing. Significant developments incorporated in the module include: (1) a sophisticated morphological analysis procedure, (2) a very efficient dictionary referencing organization, (3) powerful techniques for representing lexical generalizations, and (4) simple but effective means for morphological recombination of lexical components. 91 pp. Ref. (Author)

RM-6265/3-PR The MIND System: The Structure of the Semantic File. M. Kay, S. Y. W. Su. June 1970.
One of a series of memoranda describing the design, implementation, and use of the MIND system, this study provides a detailed description of the way in which the main fact file of the MIND information-management system is organized. The format of each type of record is given in the form of a declaration in the PL/1 programming language. A method is described for organizing long lists of items so that any desired entry on the list can be retrieved with a minimum of accesses (usually one) to the disk. Also included is a set of computer procedures which can be incorporated into any program that has cause to refer to the file. These make it easy to store and retrieve information and to relieve the programmer of the necessity of bearing all the details of the file representation constantly in mind. 70 pp. (Author)

PAPERS

P-2892 U.S.-Japan Seminar on Mechanical Translation; Summar of U.S. Contributions. D. G. Hays. March 1964.
A paper prepared for presentation at the U.S.-Japan Seminar on Mechanical Translation, 20-28 April 1964, in Tokyo. It outlines U.S. contributions to various aspects of this field of study, such as computational linguistics, linguistic theory, language descriptions, computers and programs, administrative arrangements, and publication channels among the institutions supporting research in this area. 11 pp.

P-3101 Endocentric Constructions and the Cocke Parsing Logic. J. Robinson. March 1965.
A discussion of the parsing logic devised by John Cocke and the problems encountered in its application to endocentric constructions. 23 pp. Bibliog.

P-3136 Large Files in Linguistic Computing. M. Kay. May 1965.
A description of the catalog system developed for handling data files in machine translation and other branches of computational linguistics. The catalog design for a Japanese-English dictionary is used as an illustration. 18 pp.

P-3243 The Transportation of Sentences for Information Retrieval. J. J. Robinson. December 1965.
A discussion of the difficulties to be overcome before it becomes practical to automate the retrieval of information expressed in natural text. The major linguistic task is to provide detailed, analytic, recognition grammars with transformational components adequate to deal with the complexities of the surface structures of natural sentences. 13 pp.

P-3408 Relational Data File I: Design Philosophy. M. E. Maron. July 1966.
The design philosophy and general objectives of the Relational Data File, an automatic question-answering system. The File permits factual data to be stored in a digital computer so that the machine can respond appropriately, giving answers to questions based on the stored information. (One of two related papers presented at the Third Annual National Colloquium on Information Retrieval, 1966.) 26 pp. Ref. (See also P-3411.)

P-3411 Relational Data File II: Implementation. R. E. Levien. July 1966.
The techniques used for implementation of the Relational Data File, a computer-based system for the storage, retrieval, and logical analysis of factual data. (One of two related papers presented at the Third Annual National Colloquium on Infor-

mation Retrieval, 1966.) 26 pp. Ref. (See also P-3408.)

P-3436 Computational Linguistics: Research in Progress at the RAND Corporation. D. G. Hays. August 1966.
A review of RAND's linguistic research program from its establishment in 1957. The linguistics group's work, in process and projected, is discussed in such areas as linguistic theory, descriptive studies of English and Russian, and development of basic computing tools. 9 pp.

P-3461 Processing Natural Language Text. D. G. Hays. October 1966.
A brief nontechnical overview of two applications of the computer to natural language text--for publication of scientific findings and for content analysis by social and behavioral scientists. A documentation system is outlined that starts at the author's typewriter and uses the computer as an editorial aide and as final typist; to help with typesetting; and to produce lists, files, etc. These applications are now feasible. To process language for content analysis, however, will require much further work in linguistics; present applications give promise of supporting that work. (Prepared for an NIH seminar on computational linguistics.) 7 pp.

P-3476 The Catalog: A Flexible Structure for Data Storage. T. W. Ziehe. November 1966.
An overall view of the Rand text and catalog system described in RM-4390 for processing natural language text for computer storage in large files. The system provides freedom to reorganize the files at will. Organization is based on the dependency tree. Data within any one class may be encoded in any form, disregarding the rest of the file, and the nature of the coding is indicated on the catalog map (index). Catalogs can be nested, and each catalog may be structured separately. A valuable space-saving device is the null datum, in which headings without any data fulfill their organizing role without occupying any physical space. Computer programs convert text into its internal representation. Formats have been defined for 7-bit magnetic tape and 36-bit-word core storage, and others will be added. Catalog transformation schemes are also planned for the future. (Presented at the Conference on Computers in Humanistic Research, Texas A. & M., November 1966.) 22 pp.

P-3497 Acquisition, Archiving and Interchange. D. G. Hays. December 1966.
"Acquisition" is used to mean the conversion of information from a form designed for human consumption to one designed for machine input. This can be done using a standard keyboard device such as a typewriter (perhaps slightly modified, perhaps with a variable character set). Many different kinds of printed material can be encoded in ways that retain the important typographical distinctions, if the encoding conventions used for each item are stored with the item to enable it to be printed out for reading later. The RAND Text and Catalog System for the machinable archiving of printed text is described, and a format of still greater generality is outlined. A common encoding and formatting scheme is necessary to permit interchange of materials between archives, and it must be such as to allow the publication of high-quality extracts. 24 pp. (See also RM-4390-PR, RM-4540-PR.)

P-3500 Automatic Classification in Linguistics. R. M. Needham. December 1966.
Prepared for publication in The Statistician, this Paper emphasizes the difference between methods appropriate for the classification of words and the usual forms of numerical taxonomy. Because words often have more than one meaning, linguistic classifications are overlapping. Techniques that yield a classificatory hierarchy are rankly inappropriate. The basic problem is not whether the right data have been chosen but how to handle the very large samples necessary. Randomization and construction of similarity matrices are likely to be unproductive. Independent searches for separate classes seem to work best. A nonmatrix technique employing a cohesion function and search for class profiles is outlined, which usually requires less than ten iterations for termination. 15 pp. Refs.

P-3575 Standards for Encoding Linguistic Data. M. Kay. April 1967.
Language processing is a special art that has accumulated considerable experience and technique, but few programmers are aware of it. Encoding conventions should be designed for the typist's or keypuncher's convenience. Given a simple formalized description of the input conventions used, computer programs of the RAND-Grenoble Catalog Input/Output System automatically convert the text into a standard internal coding scheme. The re-

searcher can have his output printed in one style on a local machine and then have the same material printed for permanent use with other conventions on a more sophisticated machine. The change from six- to eight-bit magnetic tape with the IBM System 360 allows for 15 different alphabets with separate upper and lower case letters, and enough codes (128) to represent most syllabaries. Global symbols, the same in all alphabets, include numerals, punctuation marks, diacritics, and indicators of italics, boldface, etc. (Prepared for publication in Computers and the Humanities.) 15 pp. Refs.

P-3600/20 Soviet Cybernetics: Recent News Items, No. 20. Edited by D. McDonald. August 1968.
This issue features a 16-page survey of Soviet computer technology, as well as a number of other informative articles: A computerized linguistic information system is being developed at the new structural typology and linguistical statistics laboratory of Moscow University. Ever-increasing production of calculating equipment still lags far behind demand, and existing computers are often inefficiently used, due partly to lack of trained personnel. A number of economic institutes do not even have punchcard equipment, and it may be necessary for them to buy equipment abroad. Each ruble invested in R&D adds 1.45 rubles to the national income, compared with only 39 kopeks gain from ordinary capital investment. Several brief reports are included on (1) conferences on automata theory, on control systems, on optimizing production and distribution, and on the mathematization of knowledge; (2) the contributions of Ukrainian economists; (3) the possibility of a time-sharing computer system; (4) designing water supply networks on the MAVR analog computer; (5) the use of computers in education and statistical linguistics; (6) automated control systems, particularly in electrical power, chemistry, and metallurgy; (7) chess and heuristic programming. 88 pp.

P-3645 Linguistic Problems of Denotation. D. G. Hays. August 1967.
The division of the linguistic properties of knowledge--speaking and understanding speech--from the psychological problem of denotation--relating perception to knowledge--is crucial to the future of linguistics. The form proposed here is to call the part of knowledge comprising properties, entities, events, and their names "dictionary" and all other knowledge "encyclopedic," separating them conceptually, but linking them substantively. 6 pp.

P-3714 Linguistic Foundations for a Theory of Content Analysis. D. G. Hays. November 1967.
A discussion of the theory that content analysis of messages is a central topic in all sciences dealing with man. Assigning properties to words and then counting the appearances of the properties is insufficient; the content of a message should be analyzed in terms of what has gone before and its relationship to what the sender and the receiver already know. The contribution of linguistics is to explicate the relationship between overt expression and content. The remarkably complicated relationships between acoustic signals and content are simplified by syntactic analysis, preferably through the use of computers. Much about language cannot be explained without grand theory, and the unexplained part is what is needed for content analysis. Linguistics can furnish a partial model of the communicator; models of cognitive processing and of attitude formation and change are also needed. Serious work on content analysis could alter the social and behavioral sciences in fundamental ways. 21 pp. Refs.

P-3719 Early Syntactic Acquisition. K. Kelley. November 1967.
A study of a child's earliest pretransformational language acquisition processes. A model is constructed based on the assumptions (1) that syntactic acquisition occurs through the testing of hypotheses reflecting the initial structure of the acquisition mechanism and the language data to which the child is exposed; (2) that semantics confirms or disconfirms grammatical hypotheses and is part of the initial structure available to the child. The model produces partial or complete analyses, since the child responds differentially to sentences according to the degree to which he understands them. The acquisition process is defined in three stages: (1) words are recognized as concrete referents; (2) syntactic development begins, separating action classes from thing classes and introducing modifiers; (3) hierarchical order appears and sentence subjects begin to be recognized. Although word comprehension is influenced by contiguous words, the child's ability to associate is not limited to physically associated words. 206 pp. Bibliog.

P-3733 Oral Literature and the Structure of Language. H. Jason. November 1967.
An application of the analytic method of structural linguistics to oral literature from all continents and many cultures. Narrative structure is conceptualized as a kind of grammar; the two basic layers are the function (a unit of action in the tale) and the move, which usually consists of three functions; connectives between moves are changes in information, state, or space. Basic moves combine to form whole tales. The main tale-roles are the hero, the donor/tester who rewards or punishes, and the donor/compensator. Since functions may be active or passive from the hero's viewpoint, the "donor" may actually obtain an item from the hero, or the hero may take it by trickery or outright theft. So far this analytic method has been applied successfully except to Oceanian tales. 22 pp.

P-3746 From Semantics to Syntax. M. Kay. December 1967.
A proposed formal system, similar to predicate calculus, for relating meanings to sentences in real languages. The system allows for paraphrase--for different sentences representing the same meaning--as is not done by present interpretive theories of semantics. The language of semantic representation cannot be a logic, since it must be based on the culture and structure of natural language and hence must include notions that are not acceptable in any logic. Self-predicability must be permitted, because a word such as "short" is itself a short word. Referents need not be real; hearers seem to interpret similarly references to real and to unreal things. (Prepared for inclusion in Recent Developments in Linguistics; expanded from a version presented at the 10th International Congress of Linguists, Bucharest, 1967.) 31 pp.

P-3756 The Identity of Noun Phrases. L. Karttunen. December 1967.
A discussion of the difficulties that arise in dealing with noun phrases according to the theory of transformational grammar. Why do people distinguish without difficulty between meanings of sentences that theory finds ambiguous? Although it is incompatible with the notion of deep structure as defined by Chomsky, there must be a semantic interpretation before the syntactic rules can operate. Very little has been written about the problem of reference in transformational grammars--it is assumed that all noun phrases refer to the real extralinguistic world. To explain the facts, we should restore some of the traditional grammatical machinery and admit that some pronouns and adjectives refer to other nodes that reside either in the same sentence or in the past discourse. (Prepared for presentation at the 42nd Annual Meeting of the Linguistic Society of America in Chicago, December 28, 1967.) 21 pp.

P-3758 A Multidimensional Approach to Oral Literature: A Proposal. H. Jason. January 1968.
A framework for a descriptive analysis of oral literature at the multidimensional level. The item of oral literature is made the center of interest, and the innate laws and outer forces that shape an oral literature and its interrelations with its context are investigated. Two units are considered: (1) The term "item of oral literature" is used to refer to the text and its performance. (2) The "repertoire of the oral literature of one society" is considered an organized unit. It is assumed that a multidimensional network of coordinates designated determinants underlies oral literature. Each item of oral literature can be measured by every determinant, i.e., it will have a "value" in terms of this determinant. Thirteen determinants that have been established are discussed. 49 pp. Refs.

P-3854 What Do Referential Indices Refer To? L. Karttunen. May 1968.
A reexamination of Chomsky's concept of a referential index that ties together noun phrases referring to the same, assumedly real-world, entity. The intricate and recalcitrant bundle of problems about noun phrases includes the distinctions of definite/indefinite, generic/nongeneric, and specific/nonspecific noun phrases; anaphoric vs. deictic noun phrases and pronouns; and restrictive vs. appositive relative clauses and modifiers. Simple English (and Finnish) sentences are examined to show that an indefinite noun phrase does not always establish the existence of an external permanent referent. Instead, the term "discourse referent" is used to indicate an entity that can be referred to by a pronoun or revived by a definite description. Even nonspecific noun phrases create discourse referents, provided they appear in a clause taken to be true. (Presented at the UCLA Linguistics Colloquium, April 26, 1968.) 29 pp. Refs. (See also P-3756.)

P-3858 Identity and "The Difference." J. J. Economos. May 1968.
A discussion of Mortimer Adler's view that man is radically different from other animals because of the human power of conceptualization shown in propositional speech. This contrasts with the materialistic Identity Thesis, which, in asserting the identity of psychological and neurophysiological processes, denies that there is such a difference. The point at issue is whether human behavior requires the assumption of nonmaterial factors for explanation. If a nonmaterial cause, such as a thought, can have a material result, such as an utterance, may not a material cause result in a nonmaterial effect, such as the production of concepts? (Presented at the Conference on Philosophical Problems of Psychology in Honolulu, March 1968.) 15 pp. Ref.

P-3869 Some Thoughts on Machine Indexing. A. E. Wessel. June 1968.
On-line computers can present a text to an editor or indexer, display the rules and categories applicable, provide immediate error- and consistency-checking, and then accept the index entries, or edited text, directly into the files. (This system has been proposed by the Reiss-Davis child mental health clinic for analysis of its case records.) But fully automatic indexing is not feasible for human-language material, since semantic and syntactic relationships cannot be formulated rigorously enough for machine application. Attempts at automated indexing have disillusioned many potential users, while concentration on developing all-machine systems may prevent the development of feasible machine aid to human indexing. When input is overwhelmingm the answer may lie in machine-categorizing of items on some pragmatic basis other than analysis of their content. The RAND/U.S. Patent Office system for machine-aided patent searches will be ready for demonstration by August or September 1968. 11 pp.

P-3871 What Makes Definite Noun Phrases Definite? L. Karttunen. June 1968.
A continuation of the analysis of referentiality in noun phrases begun in P-3756 and P-3854. There are at least three different cases to be considered. (1) Anything in the immediate environment of the speaker and hearer toward which their attention is directed becomes a discourse referent whether it is explicitly mentioned or not. (2) In every discourse there is a basic set of referents known to all the participants without special identification--such as the earth, the sky, the beach, the city--which are referred to by the definite article the. (3) A discourse referent is established without any explicit introduction when its existence is readily inferred in context: reference to "a party" introduces "the hostess"; reference to "a liner" introduces "the captain"; and references to "driving" introduce "the car" and its parts, e.g., "the hood" and "the radiator." A statement such as, "Flying to San Francisco is cheap. The fare is only $13.95," perfectly clear to the hearer, would be difficult to program in terms of the predicate calculus. 35 pp. Refs.

P-3874 The Prosodic Component: Lacuna in Transformational Theory. R. Vanderslice. 25 November 1968.
A discussion of the role of the prosodic features of American English in the framework of transformational theory. Although the Chomsky-Halle model dismisses prosodic rhythm as expressive deviation, new evidence indicates that prosody is in fact perfectly regular; moreover, it expounds grammatical referential relationships closely analogous with pronominalization or ellipsis. Accent deletion reveals identifiable structural residue and is especially pertinent. A minimal set of grammatically relevant binary features is shown to satisfactorily account for prosodic patterns: accent, emphasis, cadence, and endglide--plus a feature discriminating strong and weak syllables. Adopting these features, it is claimed, will permit explanation of more supra-sentential phenomena while positing less structure than the Chomsky-Halle stress-cycle theory which operates beyond the level of the word. 62 pp. Bibliog. (KB)

P-4023 Computational Linguistics at Rand - 1967. M. Kay. 3 February 1969.
A Paper presented at the April 1967 Rand Board of Trustees Meeting in Washington, D.C. One reason for abandoning machine translation as an immediate goal of computational linguistics was that too little was known about the algorithms and techniques that will best accomplish what linguists want to do. Consequently, researchers at Rand have turned their attention toward the need for large files of linguistic data, concentrating on developing methods for (1) storing linguistic data in a computer-usable form and (2) building up files on text in English and Russian, with statements on grammatical structure and an information-retrieval system. Among other contributions, a very general magnetic tape format and a text encoding scheme have been developed to

provide greater flexibility required by researchers. Another part of the Rand linguistics project is a general parsing program capable of handling a very sophisticated type of grammar. Grammatical theories of transformation and stratification are currently being investigated. 15 pp. (KB)

P-4093 Computational Competence and Linguistic Performance. M. Kay. May 1969.
The May 1969 presidential address to the Association for Computational Linguistics in which the activities encompassed under the term "computational linguistics" are discussed. The proof of any theory about linguistic performance is whether a machine embodying that theory can reproduce the observed phenomena. The distinguishing characteristic of the true computational linguist is that he is as much a computer scientist as he is a language expert; while the ordinary linguist tries to find the weakest formalism that will serve, the computational linguist seeks the strongest formalisms possible. His computer programs are so intimately involved with his theories that they cannot be left to a programmer to write. 18 pp. (MW)

P-4095 The Computer System to Aid the Linguistic Field Worker. M. Kay. May 1969.
A nontechnical description of the data compilation and file searching help a computer can give the practical linguist in unraveling the mysteries of exotic languages. (Flexible programs for that purpose are being constructed at the University of Chicago.) Instead of searching through boxes of paper slips, the computer-aided linguist will scan pages of machine-filed and daily-updated printout, searching for general rules and significant examples and directing the program to search for other examples and counter-examples whenever he formulates a new hypothesis. Dictionaries and concordances can be compiled and changed constantly, and the system can readily check that no material from one side of a bilingual dictionary is inadvertently left out of the other side of it. 16 pp. (MW)

P-4233 Review of The Definite Article in English Transformations. R. S. Jackendoff. October 1969.
The intent of Beverly L. Robbins' book is to confirm Henry Hiz' conjecture that, in the framework of Harris' theory of transformations, anaphoric and determinative the are entirely absent from kernel sentences and can always be introduced as transformational constants. The book develops transformations for deriving individual noun phrase constructions containing the definite article. Aside from anaphoric the the constructions all receive workmanlike treatment. No attempt is made to explain why, given the large number of transformations that produce noun phrase constructions, so many introduce determinative the as a transformational constant. The reviewer also presents three cases in which Robbins' analysis for anaphoric the cannot work. 11 pp. (MT)

P-4299 Gapping and Related Rules. R. S. Jackendoff. February 1970.
The English rule of Gapping, which accounts for such sentences as Max plays krummhorn and Boris fluegelhorn, is investigated in detail. It is contrasted with the rule VP deletion, which produces Max plays sax and Boris does too; the two rules are shown to have rather different environments. The existence of the rules N-Gapping and N-Deletion is demonstrated, and these rules are also shown to be distinct. The two N rules are then shown to generalize with Gapping and VP Deletion respectively, under the theory of syntactic nodes as matrices of distinctive features introduced in Chomsky's "Remarks on Nominalizations." 23 pp. Bibliog. (Author)

P-4391 Performance Grammars. M. Kay. June 1970.
Broad outline of a scheme to develop a machine and theory that would remove the distinctions between a speaker's semantic competence and his knowledge of the world. The structural semanticist today claims that there is no isolatable entity that can reasonably be called the meaning of a word or phrase, but that there are relationships between words and sentences that determine the meaning. To build a model of language performance, linguists must dispense with fundamental relations such as incompatibility, antonymy, and hyponomy, which are all dependent on the notion of implication. To do its job, the machine must be able to remember facts clearly and make inferences so that it can answer any question, the answer to which is not contained in any single sentence that it has stored. If the mechanisms that the machine embodies are severely restricted in some way, then the claim that it can speak and understand becomes an interesting one, and the theoretical principles underlying the

design of the machine become contributions to linguistic theory. 17 pp. (KB)

P-4481 On Some Questionable Arguments about Quantifiers and Negation. R. S. Jackendoff. October 1970.
Recent work in generative grammar, particularly in the school of generative semantics, has relied heavily on the hypothesis that quantifiers such as all and each do not originate in the noun phrases in which they occur on the surface, but rather that they are represented in the deep structure as main verbs of higher sentences. This paper shows that all the original arguments that quantifiers are higher verbs are inconclusive even within their own framework. It also questions certain proposals about the syntactic behavior of negation. 38 pp. Ref. (LC)

P-4501 A Reconsideration of Dative Movements or, Beating a Dead Horse Back to Life. R. S. Jackendoff, P. Culicover. November 1970.
A contribution to transformational grammar, explaining the indirect object shifts involved in passive versus active sentences. Noting that the preposition next to the verb is always deleted, a number of object-permuting rules are generalized into 1 rule of prepositional phrase movement and 2 rules of preposition deletion. The remaining problems of acceptability can be understood in terms of the hearer's difficulty in interpreting the resulting word strings. Contrary to the hypothesis that he reconstructs the deep structure by doing the transformations backward, it is hypothesized that the hearer first establishes known noun phrase positions, then fits NPs that have been moved into the gaps. Difficulty in finding the gap where the phrase was removed makes a sentence unacceptable. This explanation in terms of perceptual strategy accounts for the fact that 3 independent rules have identical restrictions. (Prepared for Foundations of Language.) 28 pp. Bibliog. (MW)

Bibliography: Linguistics SB-1035

ORDER FORM FOR RAND PUBLICATIONS

To: Communications Department
The RAND Corporation
1700 Main Street
Santa Monica, California 90406

Date: _____

Gentlemen:

Please send me copies of the RAND publications indicated below. I enclose payment.

(Please print)

Name: _____

Address: _____

_____

TOTAL COST: $ _____

California residents add 5% sales tax: _____

TOTAL AMOUNT ENCLOSED: $ _____

Quantity	Publication	AD Number*	Cost**	Total Cost
	R-661-PR	719 752	$2.00	
	RM-2712-AFOSR	257 286	No Charge	
	RM-2713-AFOSR	257 287	1.00	
	RM-2714-AFOSR	257 288	1.00	
	RM-2803-AFOSR	264 769	4.00	
	RM-3161-PR	276 152	1.00	
	RM-3190-PR	277 520	3.00	
	RM-3235-PR	281 850	2.00	
	RM-3339-PR	295 651	4.00	
	RM-3383-PR	288 133	4.00	
	RM-3610-1-PR		4.00	
	RM-3625-PR	405 561	1.00	

* Publications for which DDC document (AD) numbers are given are available to established users from the Defense Documentation Center (Cameron Station, Alexandria, Virginia 22314) or from the Clearinghouse for Federal Scientific and Technical Information (Springfield, Virginia 22151).

**Cost when ordering direct from Rand.

-over-

Quantity	Publication	AD Number*	Cost**	Total Cost
	RM-3774-PR	429 524	3.00	
	RM-3889-NSF	431 868	3.00	
	RM-3892-PR	432 308	1.00	
	RM-3894-3		2.00	
	RM-3979-PR	430 739	2.00	
	RM-4005-PR	432 036	3.00	
	RM-4065-PR	437 589	3.00	
	RM-4087-PR	602 648	2.00	
	RM-4136-PR	602 071	2.00	
	RM-4224-PR	605 823	2.00	
	RM-4283-PR	605 822	2.00	
	RM-4383-PR	611 306	2.00	
	RM-4390-PR	456 948	3.00	
	RM-4477-PR	626 207	5.00	
	RM-4479-PR	613 469	1.00	
	RM-4532-PR	615 319	2.00	
	RM-4540-PR	629 606	3.00	
	RM-4582-PR	620 932	3.00	
	RM-4645-PR	623 938	2.00	
	RM-4654-PR (Part I)	621 310	5.00	
	RM-4654-PR (Part II)	621 311	5.00	
	RM-4828-PR	626 572	1.00	
	RM-4920-PR	637 303	4.00	
	RM-4933-PR	635 934	3.00	
	RM-4986-PR	630 985	3.00	
	RM-5077-PR	638 924	3.00	
	RM-5085-PR	642 120	2.00	
	RM-5108-1-PR	683 724	5.00	
	RM-5143-PR	640 962	2.00	
	RM-5150-PR	652 201	2.00	
	RM-5168-PR	640 617	2.00	
	RM-5194-RADC	648 808	2.00	
	RM-5209-PR	645 451	2.00	
	RM-5210-PR	644 929	1.00	

Quantity	Publication	AD Number*	Cost**	Total Cost
	RM-5214-PR	646 598	2.00	
	RM-5223-PR	652 200	2.00	
	RM-5224-PR	646 663	2.00	
	RM-5243-RADC	649 329	2.00	
	RM-5252-RADC	649 252	2.00	
	RM-5279-PR	651 122	2.00	
	RM-5345-PR	651 406	3.00	
	RM-5452-PR	660 551	2.00	
	RM-5580-1-PR	690 990	3.00	
	RM-5611-1-ARPA	694 442	2.00	
	RM-5733-PR	673 474	3.00	
	RM-6053-PR	691 807	2.00	
	RM-6223-PR	705 544	3.00	
	RM-6265/1-PR	706 409	3.00	
	RM-6265/2-PR	709 176	3.00	
	RM-6265/3-PR	709 875	3.00	
	P-2892	433 850	1.00	
	P-3101	614 896	1.00	
	P-3136	615 301	1.00	
	P-3243	624 890	1.00	
	P-3408	636 299	2.00	
	P-3411	636 311	2.00	
	P-3436	637 929	.50	
	P-3461	640 658	.50	
	P-3476	642 368	1.00	
	P-3497	644 559	1.00	
	P-3500	644 961	1.00	
	P-3575	650 391	1.00	
	P-3600/20	673 833	3.00	
	P-3645	657 006	.50	
	P-3714	662 335	1.00	
	P-3719	662 336	5.00	
	P-3733	662 321	1.00	
	P-3746	663 215	2.00	

Quantity	Publication	AD Number*	Cost**	Total Cost
	P-3756	663 501	1.00	
	P-3758	664 604	2.00	
	P-3854	670 502	2.00	
	P-3858	671 476	1.00	
	P-3869	671 989	1.00	
	P-3871	671 478	2.00	
	P-3874	680 742	2.00	
	P-4023	684 697	1.00	
	P-4093	688 993	1.00	
	P-4095	688 859	1.00	
	P-4233	696 805	1.00	
	P-4299	701 798	1.00	
	P-4391	708 926	1.00	
	P-4481	714 330	2.00	
	P-4501	721 272	2.00	